TEC-0034

ARPA Unmanned Ground Vehicle Stereo Vision Program at Teleos Research

H. Keith Nishihara

Teleos Research
576 Middlefield Road
Palo Alto, CA 94301

April 1993

Prepared for:
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209-2308

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
Fort Belvoir, Virginia 22060-5546

US Army Corps of Engineers
Topographic Engineering Center

T E C

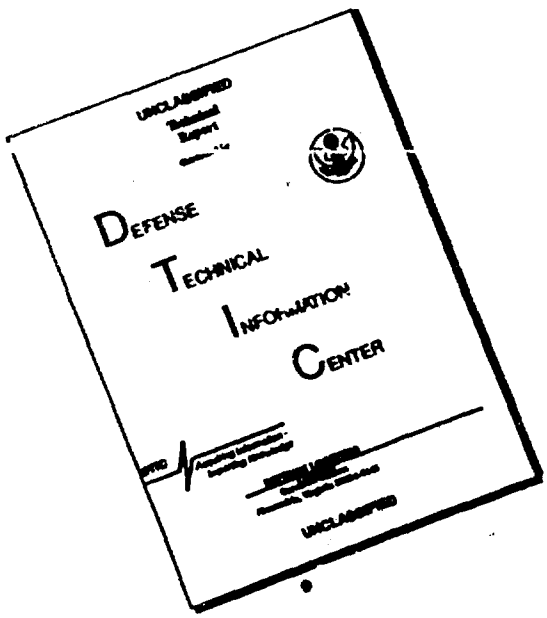
U.S. ARMY TOPOGRAPHIC ENGINEERING CENTER
TEC

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1993	Annual Report Dec. 1991 - Dec. 1992

4. TITLE AND SUBTITLE

ARPA Unmanned Ground Vehicle Stereo Vision Program at Teleos Research: Annual Report

5. FUNDING NUMBERS

DACA-76-92-C-0005

6. AUTHOR(S)

H. Keith Nishihara

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Teleos Research
576 Middlefield Road
Palo Alto, CA 94301

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Advanced Research Projects Agency
1400 Wilson Boulevard, Arlington, VA 22209-2308

U.S. Army Topographic Engineering Center
Fort Belvoir, VA 22060-5546

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

TEC-0034

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report reviews the work done during 1992 at Teleos Research in support of ARPA's UGV program. It first gives a broad overview of Teleos' approach to studying visual perception. In particular, the concept of minimal-meaningful-measurement tools is developed as a natural methodology for allowing a higher level application process to easily influence and exploit basic measurement modalities. The sign-correlation algorithm under development at Teleos is then described in this context and the current performance benchmarks of our accelerator technology are presented. Several important results in stereo theory were obtained this year including: the development of techniques for automatically setting stereo matcher operating parameters, an analysis identifying the principal parameters affecting the magnitude of the disparity gradient effects, and a technique for improving area correlator performance in the presence of large stereo disparity gradients. Contributions were also made to the development and implementation of methods for evaluating stereo matching algorithms. UGV application specific studies were carried out in the areas of narrow-field-of-view stereo; wide field-of-view, high-resolution stereo mosaic building; stereo landmark navigation; 3-D tracking of moving objects; evaluation of hardware accelerators; and development of a low-cost mobile test facility.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Stereo vision, active vision, real-time vision, unmanned ground vehicle, disparity gradients, area correlation	36
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

This research is sponsored by the Advanced Research Projects Agency (ARPA) and monitored by the U.S. Army Topographic Engineering Center (TEC) under Contract DACA76-92-C-0005, titled "ARPA Unmanned Ground Vehicle Stereo Vision Program at Teleos Research". The ARPA Program Manager is LTC Erik Mettala, and the TEC Contracting Officer's Representative is Ms. Linda Graff.

1 Introduction

This report reviews the work done during 1992 at Teleos Research in support of DARPA's UGV program. Our research activities have been divided between basic studies relating to binocular stereo and technology development relevant to the UGV mission. Highlights of this years work include the development of several new algorithms for enhancing stereo matcher performance on UGV relevant imagery. These include (1) techniques for automatically setting stereo matcher operating parameters such as filter size by previewing results on a sparse set of points over a range of possible parameter settings; (2) an analysis identifying the principal parameters affecting the magnitude of the disparity gradient effect that compromises correlator performance in UGV stereo imagery; and (3) a technique for improving area correlator performance in the presence of large stereo disparity gradients.

Teleos also collaborated with the other UGV stereo contractors on a broad evaluation of stereo matching algorithms. A suite of stereo imagery for testing matcher performance in the presence of increasing noise and in the presence of large disparity gradients was contributed to the project. Teleos also submitted its own stereo matching algorithm to the evaluation process and that algorithm performed well overall and was noteworthy among all compared for its noise handling capabilities.

During this period Teleos carried out UGV related studies in the areas of narrow-field-of-view stereo: wide field-of-view, high-resolution stereo mosaic building; the feasibility of using stereo landmarks in support of vehicle navigation; active control of a stereo sensor head for 3-D tracking moving objects: feasibility of porting our matching algorithms to several types of parallel processors; and development of a vehicle based test facility for supporting the study of real-time stereo using an active sensor head.

This report first presents a review of some of the principles guiding Teleos' research in real-time stereo and motion perception. The report then reviews the core research done under the program. Then the UGV specific technology development done in support of the contract is discussed.

2 Task directed visual perception

One of the goals of our research has been to develop and understand practical demand-driven computer vision. Our view is that the sophisticated performance observed in biological systems is, to a large degree, derived from the fluent use of simple and robust measurement capabilities. We are attempting to identify and study early modular perceptual abilities in biological systems that fit this model, and we have worked extensively with two: stereo disparity and optical flow measurement.

Research on stereo and motion sensing and enabling processing technology has matured to the point where interesting real-time applications such as vehicle guidance and security and surveillance are practical. Three broad questions are pertinent to the design of visual perception capabilities supporting applications like these: (1) what specifically should we measure, (2) what are the best algorithms to use, and (3) what is the best hardware technology to build on?

2.1 Minimal Meaningful Measurement Tools

The perceptual information a blind person needs about his environment and the character of aids that prove most useful to him can provide practical guides for research in machine vision. There is a close analogy between the sensing needs of an intelligent blind person and those of active problem solving machines.

To be acceptable for use by a blind person, a visual aid must be easy to use, informative, and cost-effective. Interestingly, aids that are "too smart" are often rejected because they leave the blind user oblivious to much of the detail of what is going on. This makes it hard for that person to use the tool effectively in new contexts. What seems to be called for, in the case of the blind, are aids that such users can operate as tools to accomplish perceptual tasks.

Following this line of thought, a desirable perceptual aid for machine vision ought to recover some basic information and it should have an easy-to-model behavior that is sufficiently rich to allow an expert to use it in creative ways. A blind person's cane is a good example: it has a consistent mechanical behavior and it provides timely information about the presence or absence of physical objects at dynamically selected locations about the

operator. The cane "device" has low-bandwidth input and output interfaces to the user—that is, manual pointing control and force, vibration, and sound feedback. This allows it to be managed easily by the blind user while carrying on other parallel activities such as conversation. Furthermore, though simple, the cane has a fairly rich and consistent behavior that fosters the development of expertise in its use. For example, one learns the *feel* of different pavement textures or conditions—slippery or uneven.

We think of a measurement tool as a device that a higher level agent can deploy "skillfully" in specific task domains, much as a blind person uses a cane or as an artist uses a brush. Three qualities are noteworthy of such tools:

1. **simple but meaningful.** The device should make the simplest meaningful measurement possible to be efficient. Too much automatic interpretation at this level can be counterproductive and too many gratuitous measurements can waste processing resources. This orientation makes it is easier to present more precise information to the user and it allows the user to interpret the basic measurements with increased efficiency and precision.

2. **easy-to-model.** The device should have a consistent, easy-to-model behavior. If the underlying algorithm has many special case behaviors, it becomes difficult for a user to anticipate that device's behavior in new situations or possibly even in familiar ones.

3. **informative output.** The device should exhibit a behavioral richness that encourages the learning of strategies for making more specialized measurements with it. For example, simply reporting best estimates of range from a stereo correlation tool would deprive the user of valuable information about the shape of the correlation peak. In various circumstances, that user might be able to use knowledge of the peak's height, its broadness in vertical disparity, or its bimodality.

This measurement tool concept can be applied to the study of early vision problems to help us define computational problems that are somewhat different from the problems that are traditionally addressed. For example, instead of attempting to compute a dense stereo range map, we concentrate on the

problem of computing and communicating the results of a single range measurement over a patch of surface. This distinction can be significant when issues of interaction with higher level knowledge and control are considered.

In stereo matching, a measurement over a small sensing area may fail due to the absence of matchable features. To recover, the calling agent can try switching to a larger measurement window or it can move the orignal measurement patch to a slightly different position, or it might decide to move the sensor head to a better vantage position. In either case, the calling agent is aware of the changes made and their implications for the measurement. It is in possession of knowledge of the task to be accomplished, it is aware of the measurement difficulty and the character of the possibly degraded information obtained. At the same time this agent does not have to know much about the detailed workings of the measurement algorithm itself. As long as it exhibits a consistent and predictable behavior it can be used effectively treated as a black box.

2.2 Sign-Correlation Image Matching Theory

The first class of computations studied extensively in this context have been image matching algorithms applicable to stereo range finding and optical flow field measurement. We have developed a computational theory for measuring stereo and motion disparity that is consistent with the measurement tool objectives and we have had some success at demonstrating the validity of that model for biological systems. We have also developed practical real-time hardware accelerators for our algorithms.

Binocular stereo, the measurement of optical flow, and many alignment tasks involve the measurement of local translation disparities between images. Marr and Poggio's zero-crossing theory made an important contribution towards solving this disparity measurement problem[1]. The zero-crossing theory, however, does not perform well in the presence of moderately large noise levels as has been illustrated by the inability of zero-crossing based approaches to solve transparent random-dot stereograms--which, interestingly, can be perceived correctly by the human visual system[2]. We have since developed a sign-correlation algorithm that builds on Marr and Poggio's ideas and that addresses many of the weaknesses of the original work.

We continue to use the zero-crossing primitive for matching, but the matching rule is changed. Instead of matching zero contours, we correlate

the signal's *sign* in an area. This subtle change makes a significant difference in the behavior of the matcher. Sign-correlation continues to provide useful disparity measurements in high noise situations long after the zero-crossing boundaries, surrounding the signed regions, cease to have any similarity. An intuitive explanation of why the two approaches perform so differently follows from the fact that the sign of the convolution signal is preserved near its peaks and valleys long after increasing noise has caused the zero contours to be fully scrambled. Thus area correlation of the sign representation yields significant correlation peaks even with signal-to-noise ratios of 1 to 1. Since sign-correlation still operates off of the zero crossing representation, the key strengths of Marr and Poggio's theory are preserved.

The high noise tolerance of our matching algorithms makes them particularly well suited for use with night vision equipment, such as intensified cameras, which exhibit high shot noise levels, and IR cameras, which have low contrast levels as well.

2.3 Acceleration for real-time operation

We have made significant advances in developing algorithmic and hardware techniques for accelerating the large kernel convolutions and area correlations used by the sign-correlation approach. At present a pair of VME bus boards along with a general purpose processor board carry out full frame stereo convolutions at video rate with $\nabla^2 G$ convolution operators as large as 60 by 60 pixels. Stereo disparity measurements covering a search space of 72 by 4 pixels at a typical resolution of .2 pixels are accomplished in 300 microseconds.

Area-based motion measurement can be done using the same facility with an interframe search range of 108 by 108 pixels and the same subpixel resolution, in under 2 milliseconds. This allows the tracking of bodies moving at angular rates of 30 degrees per second with a sensor having a 10 degree field of view. That translates to being able to detect and follow a subject entering the field of view traveling at 36 km/h at a range of 10 meters.

Optical flow fields and stereo range measured sparsely over the entire visual field can be used to do rapid figure ground discrimination and this result can then be used to focus attention and further processing on meaningful physical entities.

3 Stereo core research

During the program year we completed four tasks aimed at improving the performance of a stereo matching system such as our sign-correlation system when operating on UGV-relevant imagery. These efforts resulted in (1) an automated technique for selecting pre-processing filter sizes appropriate for dynamically varying scene characteristics; (2) an analysis of a disparity gradient effect which can have a significant effect on stereo matcher performance in UGV imaging configurations; (3) a technique for efficiently mitigating the disparity gradient effect; and (4) an evaluation of our matching algorithms carried out in cooperation with other UGV stereo team members.

3.1 Control parameter selection

The principal control parameter of the sign-correlation approach is the size of the convolution filter used. This parameter selects the spatial scale at which texture is picked up for use in the stereo correlation. In many cases a very large operator accentuates texture that is more stable than that available at finer scales. Since the correlation window size scales with the filter size, it is desirable to find the smallest filter size that yields acceptable correlation measurements.

To automate the selection of an appropriate operator size we developed an algorithm that pre-samples the stereo image at a small number of locations and at each of these it checks the quality of the stereo correlation peak over a range of filter sizes.

Specifically the algorithm does a search over filter size, w, from the set: {4, 6, 8, 12, 16} (units are pixels). For each filter size, correlation statistics are sampled at 25 locations evenly spaced over the filtered stereo pair.

At each of the 25 sample locations, we search for the highest correlation peak in the disparity search range. We also keep track of the second highest peak. For each of these two peaks we compute the peak's *local height* above the correlation values at a distance in horizontal disparity of $.75w$ from the peak disparity. This prevents blank regions in the images from looking good. We then take the difference between the local height of the best peak and the second best peak at the sample location. This *peak-difference* allows us to fold in a requirement that there not be multiple peaks at the sample location.

We select the smallest filter satisfying at least two of the following three criteria:

1. Average peak-difference score is above 0.3.
2. At least 90 percent of samples have peak-differences better than 0.1.
3. At least 80 percent of samples have peak-differences better than 0.25.

If no filter size satisfies this test, the filter with the largest average peak-difference is used.

This algorithm exhibits the following characteristics:

1. It favors smaller filter sizes which yield better spatial resolution. If all else is equal, we do better with smaller filter sizes which allow us to use correspondingly smaller correlation window sizes.
2. It enlarges filter size when shot noise levels are high. Shot noise is prevalent in low contrast areas of an image and in many night vision sensors. When these levels get sufficiently large relative to the local image texture contrast, matching performance drops. A larger filter often will increase the relative strength of coarse texture patterns relative to this type of noise.
3. It enlarges filter size to avoid ambiguous correlation peaks. In some imagery there is repeating structure, such as checkerboard patterns, that cause ambiguous correlation peaks. These repeating patterns are sometimes not present at coarser scales because they are dominated by other coarse scale variations, such as slight irregularities in the checkerboard.

This filter size preselection technique operated effectively on the large JISCT evaluation project discussed in a later section of this report.

Figure 1 shows a stereo pair prepared by CMU to illustrate the ambiguity problem that binocular stereo matchers can run into. The shoe appears to be sitting on a rubber door mat but is actually held several inches above the mat. The repetitive pattern present in the mat allows most matchers to incorrectly follow the false correlation surface that continues from the shoe's sole out over the door mat. CMU researchers have argued that solving this type of problem requires the use of multiple baseline stereo.

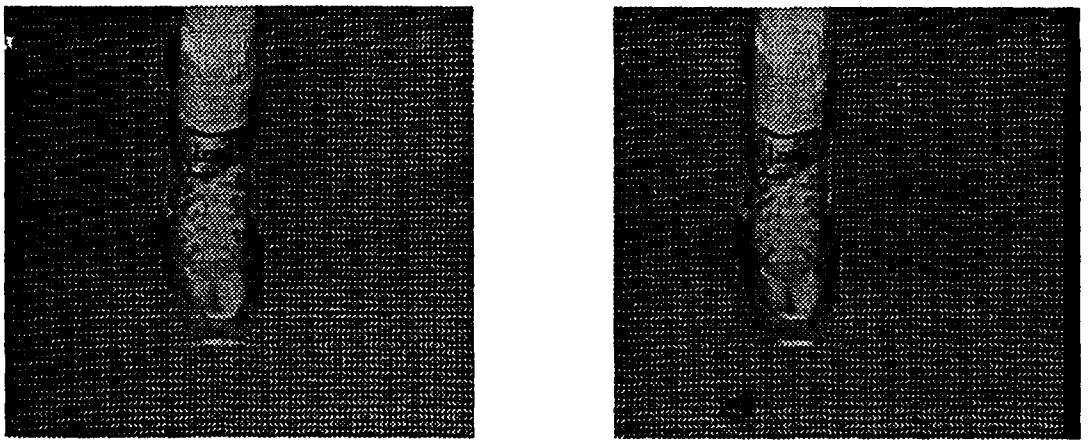


Figure 1: CMU Shoe stereogram illustrating ambiguity problem for binocular stereo. The shoe is actually held several inches above the door mat below it, but this is not easy to detect because of the regular texture on the mat.

Figure 2 shows the sign representation obtained from figure 1 when filtered with a medium sized $\nabla^2 G$ operator. As one would expect, the repeating pattern on the door mat shows up crisply and the sign correlation surface on the mat exhibits multiple ambiguous peaks. Interestingly, when our filter size selection algorithm was applied to the CMU shoe stereogram, it selected a much larger filter size which yielded the sign representation shown in figure 3.

With this larger filter the repeating pattern of the door mat has been replaced by a coarser texture pattern associated with the pattern of irregularities in the mat. Thus unambiguous disparity measurements can still be made using binocular images with the larger filter size. Figure 4 shows a disparity surface plot computed using the sign correlation algorithm. There is some rounding of the surface at the shoe edges due to the large correlation windows used. This averaging effect can be reduced by following the coarse matching step with subsequent passes using smaller filter sizes using the coarse data to disambiguate the multiple correlation peaks.

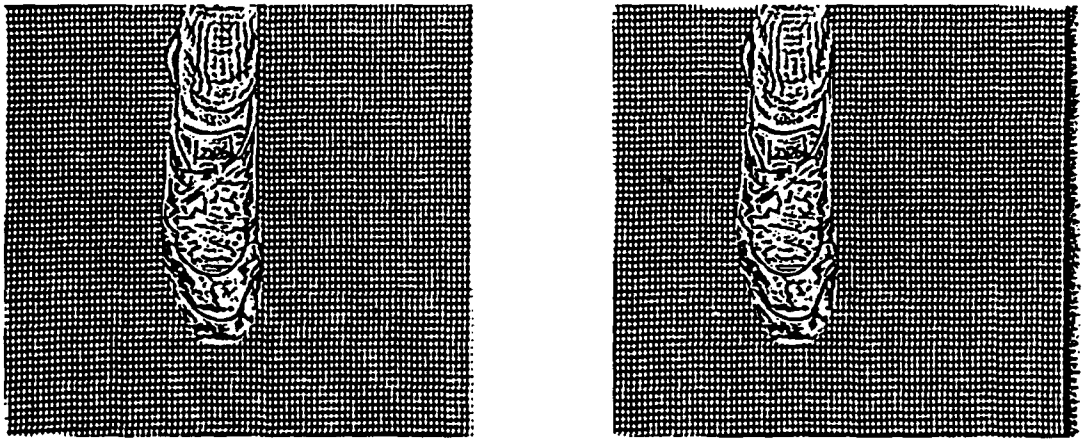


Figure 2: Sign representation of figure 1 using a medium sized filter. The repeating texture on the door mat is apparent and this makes matching ambiguous.

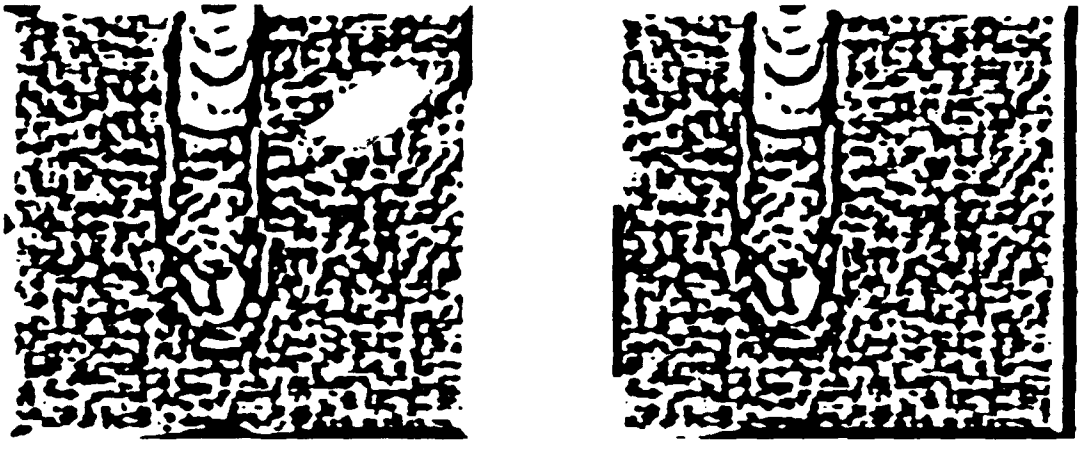


Figure 3: Sign representation of figure 1 using a filter 3 times larger than that used in figure 2. Note that the repeating texture is no longer present. Instead we see coarse textures associated with irregularities and dirt on the door mat.

Figure 4: Disparity surface plot of the CMU shoe stereogram computed using the sign representation shown in figure 3.

3.2 The disparity gradient effect

We obtained a new and rather surprising result regarding stereo disparity gradients in the UGV imaging configuration, where the stereo cameras are mounted on a vehicle looking out ahead at the ground. These gradients occur when the cameras view an inclined surface such as the flat road out in front of the vehicle as depicted by figures 5 and 6. They can significantly affect the performance of area correlation based matchers because the receding surface under a correlation window does not register at any single disparity. This causes the correlation peak obtained to be lower and spread out making detection of the peak more difficult and unstable.

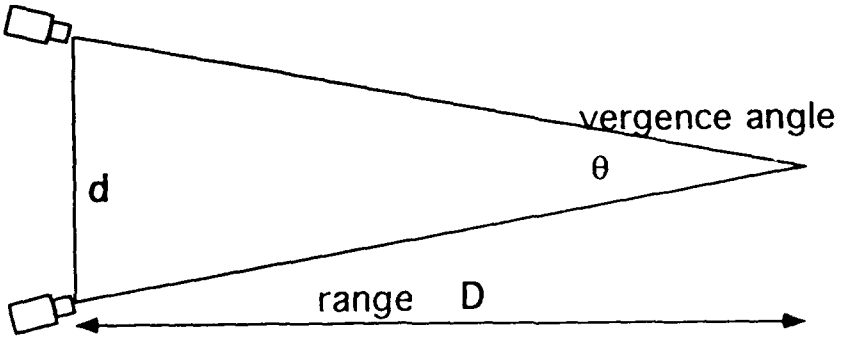


Figure 5: Aerial view of stereo imaging configuration.

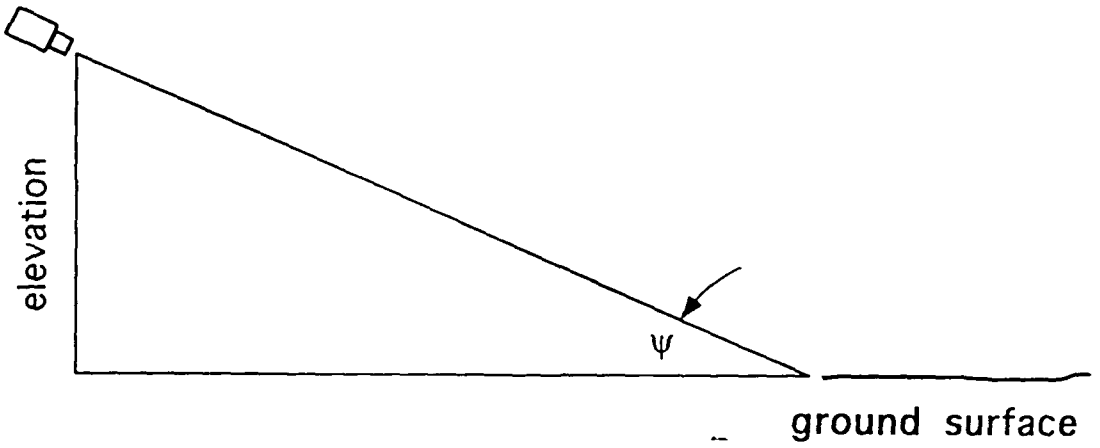


Figure 6: Side view of stereo imaging configuration.

We were able to derive an estimate for the disparity gradient magnitude as a function of stereo sense head parameters such as camera lens size, baseline separation between the cameras, camera height, and the pitch angle of the cameras (how much down from the horizon are we looking). The expression obtained is:

$$disparity\ gradient \approx \frac{baseline}{height} \tag{1}$$

In other words, the disparity gradient depends primarily on the ratio of camera separation to camera height. It does not depend significantly on lens size, or pitch angle of the cameras so long as the cameras are looking significantly farther ahead than they are high above the ground.

An important consequence of this result for the UGV program is the constraint it imposes on the baseline separation. Typical matching algorithms are seriously affected by gradients larger than about 0.2 pixels disparity change per pixel in the image. This means that camera separation should not be larger than one fifth of the height of the camera head above the ground.

We derive the approximation of equation 1 using the notation on figure

7 as follows:

$$\frac{elevation}{D} \approx \frac{\Delta s}{\Delta D} \tag{2}$$

$$\approx \frac{\Delta \alpha D}{\Delta D} \tag{3}$$

Equation 2 is by similar triangles, assuming that Δs is much smaller than D and that Ψ is small.

Rearranging terms allows us to write:

$$\frac{\Delta D}{D} \approx \frac{\Delta \alpha D}{elevation} \tag{4}$$

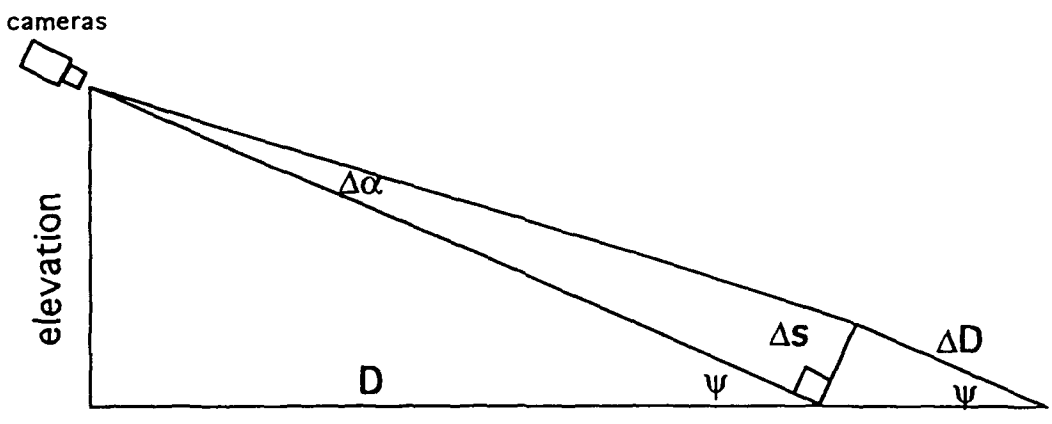


Figure 7: Relation of vertical disparity to vertical image position over a flat inclined surface at a distance.

Now from other work [3] we have the relation:

$$\frac{\Delta \theta}{\theta} \approx \frac{\Delta D}{D} \tag{5}$$

Combining with equation 4 we get:

$$\frac{\Delta \theta}{\theta} \approx \frac{\Delta \alpha D}{elevation} \tag{6}$$

rearranging and substituting $\frac{baseline}{D}$ for θ we get:

$$\frac{\Delta\theta}{\Delta\alpha} \approx \frac{\theta D}{elevation} \tag{7}$$

$$\approx \frac{baseline}{elevation} \tag{8}$$

which is the desired relation.

We have communicated this information to Martin Marietta and are working with them to see that these considerations are reflected in their sensor head design. As noted in previous reports, we are continuing to work on methods for extending the operating envelope of our matching algorithm to handle larger disparity gradients.

3.3 Skewed correlation window technique

We can compensate for vertical disparity gradients when we make a correlation measurement by progressively shifting the horizontal disparity between the left and right correlation windows as we scan vertically over those windows as shown in figure 8. Adjusting the correlation window "skew" can greatly improve the correlation peak height obtained in images with large vertical disparity gradients. The window skew that gives the best correlation can also be used to directly estimate the local disparity gradient. A similar operation can be done to compensate for horizontal disparity gradients. A future firmware upgrade to the PRISM-3 hardware accelerator will allow dynamic compensation for both vertical and horizontal disparity gradients.

Figure 12 shows one example: a stereo disparity map computed on a surface with a large vertical disparity gradient. Skewed correlation windows were used and yielded almost a doubling of correlation peak height on the examples in that test suite.

3.4 Performance Evaluation

A critical component of the stereo research and development program for the UGV mission is the definition of evaluation metrics and tests. As one of our first projects on the program, we collaborated with the other UGV stereo contractors to formulate test protocols and collect test data. We assembled

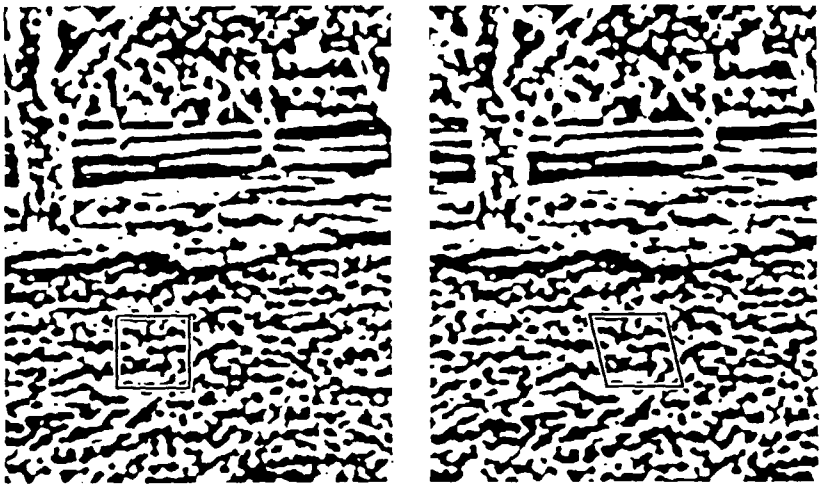


Figure 8: Correlation peak height can be substantially improved on images of surfaces sloping away from the cameras by skewing one window relative to the other as shown above. Note that the sign blob patterns in both windows match quite closely. This is not the case when rectangular windows are used on both images.

a large test database of stereo imagery, known as the JISCT database, and ran several of our best matching algorithms on them for a comparative analysis. The test stereo pairs were distributed to participating groups in early January and results using algorithms operating at SRI, Teleos, and INRIA were collected for analysis.

In support of this evaluation effort, Teleos developed a set of stereo test images designed to test performance on scenes with large vertical disparity gradients and high noise levels. Both factors are typical of UGV imagery. Our suite of test images are all of the same scene, a flat steeply inclined surface (68 degree incline) with a spherical object at the center of view. The first stereo pair of the set, shown in figure 9, is with good exposure settings on both cameras. Figure 10 superimposes graphs of the intensity profile along a horizontal raster line through the tennis ball in the left and right images of figure 9. The ball is slightly shaded and gives rise to the abrupt dip in both curves near the center of the plot. With scrutiny one can observe the correlation between intensity fluctuations in the two superimposed graphs.

Figure 11 shows the Laplacian-of-Gaussian sign patterns computed from the stereo pair in figure 9. The filter size—automatically selected—had a center diameter of 6 pixels. Figure 12 plots the disparity surface computed

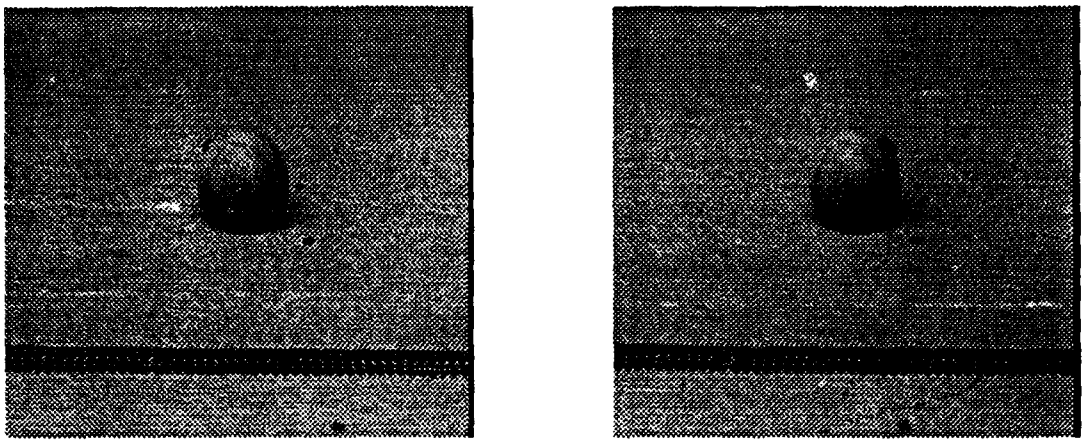


Figure 9: Stereo pair of a flat board at 68 degree incline to camera axis with a tennis ball at the center. This is first image pair of a set with increasing noise levels which tests a matcher's noise handling ability and ability to handle large disparity gradients.

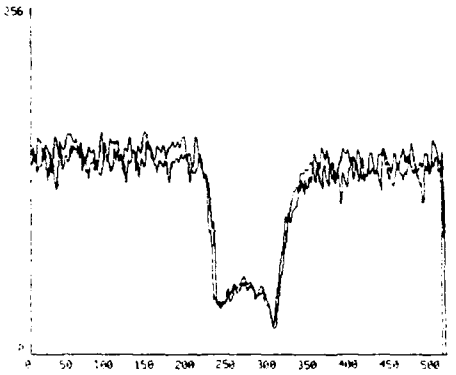


Figure 10: This graph plots the camera intensity along a raster line passing through the center of the tennis ball in figure 9. Graphs from both the left and right images are superimposed and a reasonable correlation can be seen especially over the ball. The stereo disparity between the images along this raster line is close to zero.

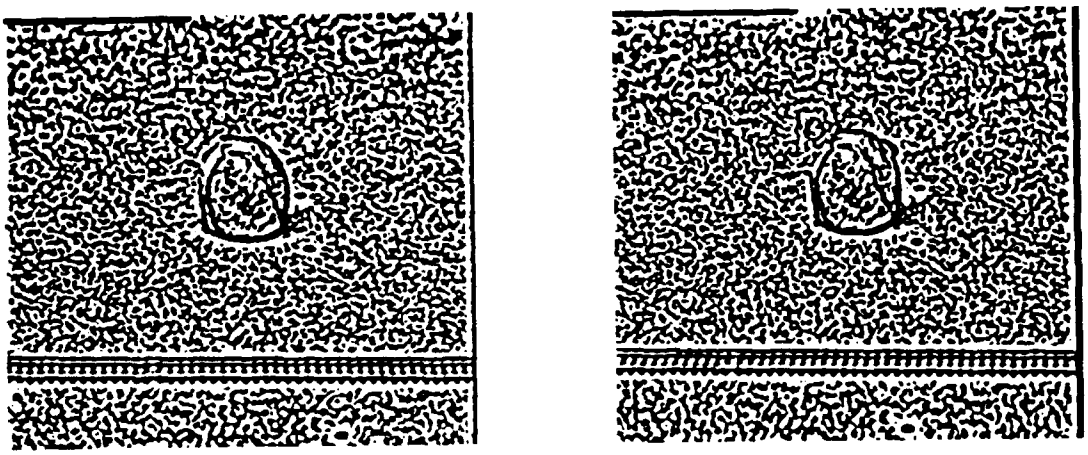


Figure 11: Sign representation of figure 9

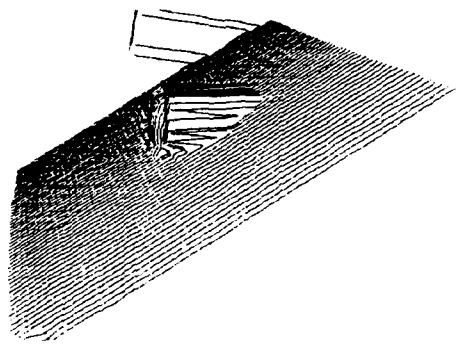


Figure 12: Disparity plot computed using the sign correlation algorithm from the sign representation shown in figure 11. Cameras were to the left and the line of sight is horizontal in the display. Note the flatness of the surface around the ball at the center. Unmatched regions—due to unsatisfactory correlation peak height or shape—were mapped to zero disparity mostly around the perimeter of the ball and in a small patch at the back edge. Nintyseven percent of the image was matched.

from correlating the sign patterns. Matches were found at all but 3 percent of the points checked and the flat plane of the surface and the curvature over the tennis ball's surface are apparent. The unmatched locations were restricted to the neighborhood of the ball's occluding boundary and to a small patch near one edge of the image.

Each successive pair in the test suite of seven stereo pairs has the camera aperture reduced by one stop on the lens aperature. This created a series of stereo images with the same stereo scene but increasing noise levels due to the decreasing amount of light available to the cameras. The last pair in the set has no useful signal present. As we expected, most stereo algorithms were able to handle the first pair of this set, but performance of individual algorithms fell off at different points in the series of test pairs.

Figures 13 through 16 show the same sequence of displays for the fifth pair in the test suite. Figure 13 shows the raw stereo pair, but here the lenses have been closed down 4 stops and we see an essentially black display. There is still a small amount of contrast remaining, as can be seen from the graphs in figure 14, but the ratio of signal to sensor and digitization noise is decreased by a factor of 16. Figure 15 shows the sign patterns obtained from the raw stereo images. In this case a slightly larger *filter center diameter* of 8 pixels was automatically selected for use in the correlator. Note that moderately stable patterns are discernible even at this increased noise level. Figure 16 plots the disparity array computed in this case. Ninety percent of the attempted points yielded acceptable correlation peaks and as can be seen from the plot most of the board surface was recovered. Similarly, the ball's curvature is still apparent. Increased dropouts occured at the edges of the image and around the ball's occluding contour. By contrast, no other algorithm participating in the JISCT evaluation was able to get any meaningful results on this stereo pair.

For the next image pair after the one shown in figure 13, with 50 percent less light, the sign-correlation algorithm still yielded matches at half of the image locations.

As noted earlier, the sign-correlation algorithm was the only one to perform well on the repeating background of CMU's *shoe* stereograms. As we saw in figures 1 through 4 these stereo pairs were intended to be an example of a matching problem that could only be solved by multiple baseline approaches. We found, however, that the coarse-to-fine scale-space filtering available with our sign-correlation approach allows the problem to be

Figure 13: This stereo pair is the fifth member of the test suite. It is of the same scene as that shown in figure 9 but with the cameras' irises have been closed down 4 stops which effectively leaves the black images shown here. There, however, is still a very low contrast signal embedded in the background noise of the cameras.

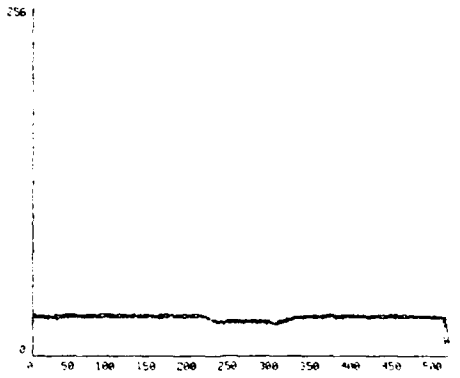


Figure 14: This graph plots the camera intensity for figure 13 at the same position as in figure 10. Note that the signal contrast is significantly reduced and it is no strong correlation between the two superimposed traces other than for the slight dip in brightness over the tennis ball.

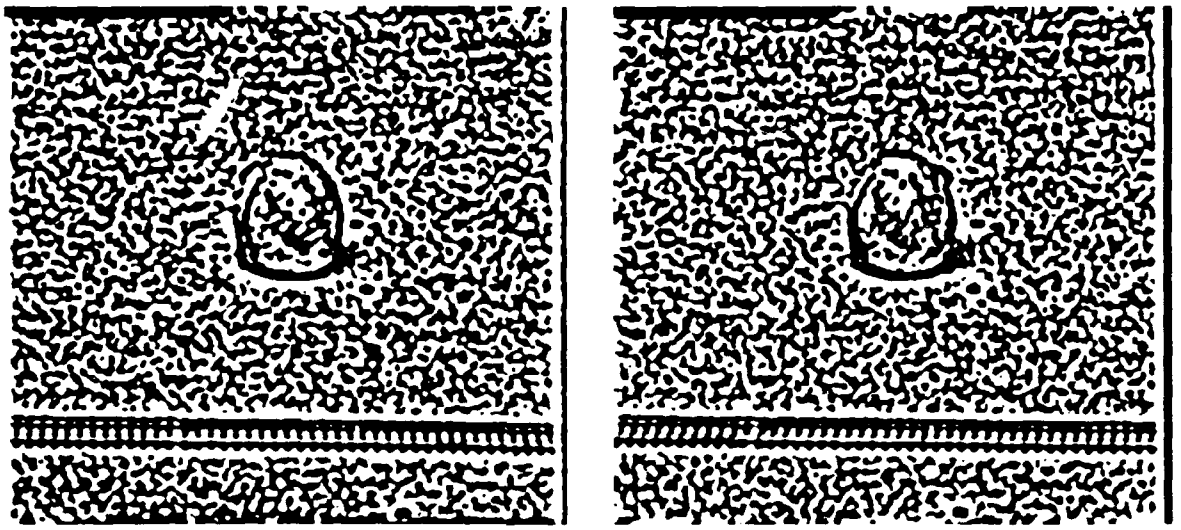


Figure 15: Sign representation of figure 13. Despite the thirty fold loss of contrast, the sign patterns remain moderately well correlated.

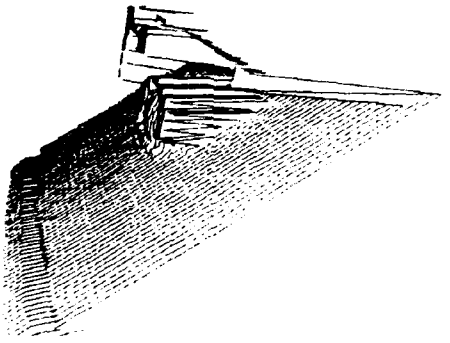


Figure 16: The disparity plot computed using the sign correlation algorithm applied to the sign representation shown in figure 15. The results are essentially the same as those shown in figure 11 with more unmatched points—showing as zero disparity—along the edge of the board surface farthest from the cameras. There is also a larger unmatched boarder around the tennis ball and the train tracks at the near edge of the board were not matched. Overall 90 percent of the image points were matched in this test

solved with a single stereo pair with any baseline. Sign-correlation succeeds in this case because the wallpaper illusion in the example is not perfect, there are small variations due to dirt or other irregularities in the floor mat that was used for the background of the scene and that introduces coarse scale textures that can be used by our matcher to make unambiguous disparity measurements.

A comparative analysis of the test results by SRI indicated that our sign-correlation approach performs very well on ground terrain and in large noise situations, and as expected, slightly less well than some other approaches at discerning small objects due to the larger filter and window sizes employed.

4 UGV technology development

During the program year, we carried out nine distinct tasks in direct support of the UGV program. These include: (1) taking the lead for investigating the possible role of narrow-field-of-view (NFOV) stereo on the UGV; (2) investigating the use of NFOV stereo to produce high resolution wide field range maps; (3) conducting feasibility study for using stereo derived landmarks to support navigation; (4) investigating the feasibility of using the IWARP processor for our algorithms; (5) contributing to stereo video data collection efforts; (6) developing a car mounted test facility; (7) experimenting with active head control for motion and stereo tracking; (8) providing technical information transfer and assistance to Martin Marietta; and (9) proposing a method for enhancing operator and observer awareness of system operation.

4.1 Narrow field of view stereo

One of our primary focuses at Teleos has been on active visual perception and we have taken the lead on the UGV program for exploring ways to apply active visual perception in support of meeting UGV mission requirements. In particular, we have identified the following tasks:

1. Far-look-ahead obstacle detection using stereo. As illustrated in figure 17, a narrow field of view stereo sensor can be configured to look for hazards out toward the horizon along the expected direction of vehicle travel.

2. Active following of navigational features. For example, when driving on a road with a steep embankment to one side, a narrow-field-of-view (NFOV) stereo sensor can be used to monitor the position of that embankment.

3. Double checking wide-field-of-view (WFOV) data. The higher resolution pixel data on a NFOV system can be used to check potential hazards detected by the WFOV system.

4. WFOV high resolution mosaics. An active system with narrow lenses can be scanned over a scene to build up a high resolution stereo range image.

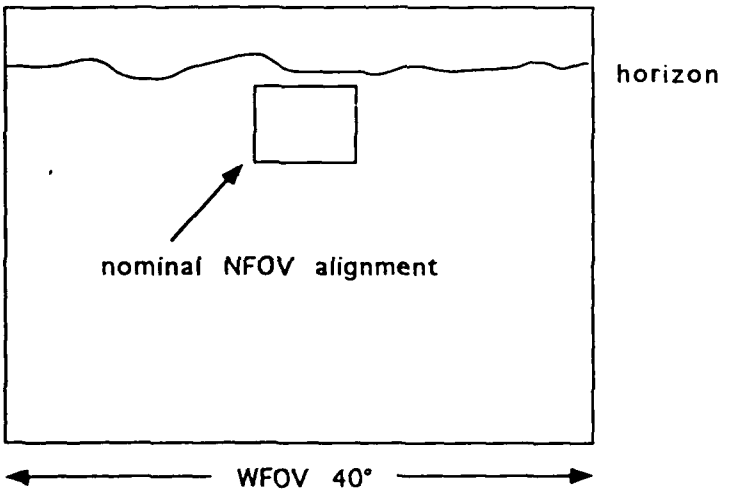


Figure 17: This figure depicts the superposition of wide field and narrow field stereo sensors to provide wide coverage and far lookahead in a more restricted zone in the direction of travel.

4.2 Wide-field high-resolution stereo

The first of the above areas to be investigated carefully [illegible] on the building of high resolution stereo range mosaics by scanning a [illegible] NFOV sensor over a large scene. Figure 18 shows the kind of result [illegible] can be obtained. In this figure, the stereo head is looking out a window at an outdoor scene with trees, buildings and an intersection. The upper image in

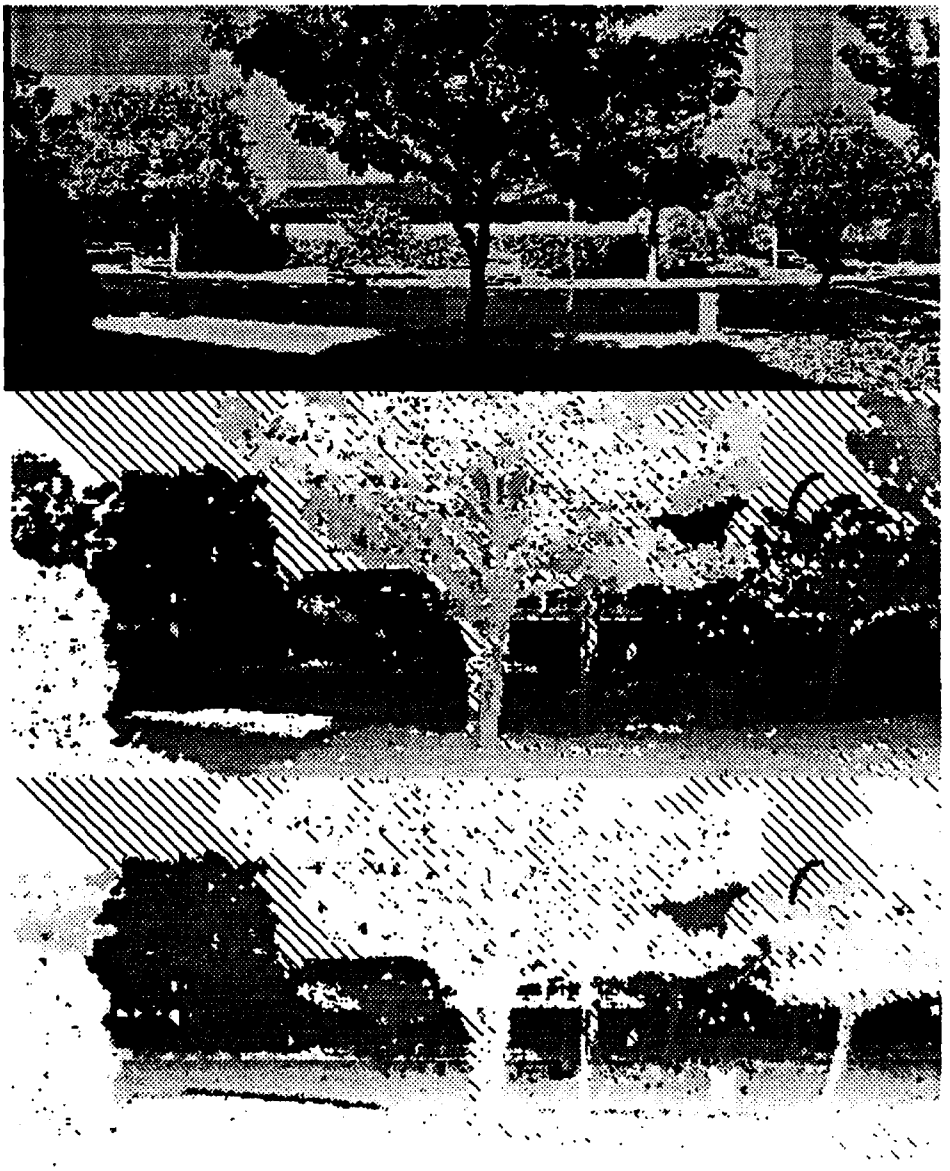


Figure 18: The upper display here shows a mosaic of many camera images of a street scene covering about 100 degrees in pan angle and 30 degrees in tilt. The lower displays show the corresponding stereo range map computed from this scene. Brighter shades of gray indicate locations closer to the camera. The bottom display shows the same range information as the middle one, but with shading set to highlight areas farther away. The hatched areas indicate unmatched regions which in this case are largely on the sky which was cloudless.

the figure shows the gray level camera image and the lower images show stereo range using shades of gray with lighter indicating closer to the cameras. The hatched areas mark locations where the matcher did not find a satisfactory correlation peak. These are largely on the sky which was cloudless. The distances of the close objects (light grey tree crowns) vary from approximately 10 to 30 feet; the distances of the far background (dark grey buildings and trees) range from 70 to 150 feet.

The sample interval in both dimensions is .25 degrees and the spatial resolution of the range measurements is about .5 degrees, or about one foot at one hundred feet. It is interesting to note that objects like the cobra light pole at the upper right are more discernible in the range image than in the raw camera image.

4.3 Stereo landmarks

An active stereo range sensor equipped with narrow field of view lenses can be used to acquire high-resolution wide field of view range data from the environment around a vehicle. This capability presents the opportunity to track movements relative to its environment, and thus assist local navigation[4]. It may also be feasible to use occupancy maps built from this data to recognize and navigate through previously traversed locales.

Potential methods for tracking a vehicle's movement through a locale based on range data were reviewed. We then implemented a voting-based system and obtained promising results when this method was applied to real outdoor stereo data.

The current system could be readily extended to a more general method of navigation. In addition to the localization of the robot within a locale, the present system could be used to identify the current locale of the robot from a database of multiple possible locales. The range map of each distinct locale could be matched to the range map measured at the current position, and the height and sharpness of the peak of each match could be used to determine the correct one. This locale recognition capability could be especially useful in discovering path cross-over points during an extended journey. Maintenance of a database of previously seen locales would also be valuable for guiding a robot back to its point of origin or to any other point along the path it has traversed.

Intelligent feature extraction from the range data, possibly augmented

with other sensed information such as color, could potentially increase the performance of the system significantly. We chose, however, to leave this for later research because the task of defining and actually recovering stable landmark features from range images, like that shown in figure 19, appeared to be beyond the scope of our initial project.

4.4 IWARP port analysis

While the IWARP processor was under consideration for use in the UGV program, we developed test code to assess the feasibility of porting our sign-correlation algorithm to that computation engine. We determined that our pipelined Laplacian-of-Gaussian hardware design could be mapped to the IWARP in a relatively straightforward manner. Timing benchmarks run on the IWARP at CMU indicated that a 64 cell IWARP could do the convolutions at about 1/3 the speed of our video rate convolver board. Analysis of the correlation stage of our algorithm showed that a 64 cell IWARP implementation would run significantly faster than our current single board correlator.

4.5 Test data recording

A data recording capability sufficient to capture live stereo video on a moving vehicle along with associated navigation and camera attitude data is on the critical path for our research effort. A high performance recording system of this type will take some time to be designed and installed. In the mean time we explored interim methods for recording stereo imagery that might be implemented quickly and inexpensively at all sites so that early test imagery can be collected and shared easily by team members.

Several approaches to solving the recording problem were formulated. The first of these involves collecting stereo data by interleaving left and right camera images on the even and odd fields of a single interlaced video frame. This sacrifices half the vertical resolution but should be sufficient to drive early performance studies. To be usable, it will be necessary to acquire the left and right half fields simultaneously while sending them out to the recorder sequentially. Likewise, it will be necessary to undo that encoding on playback for real-time processing experiments. We have implemented this approach on our system.

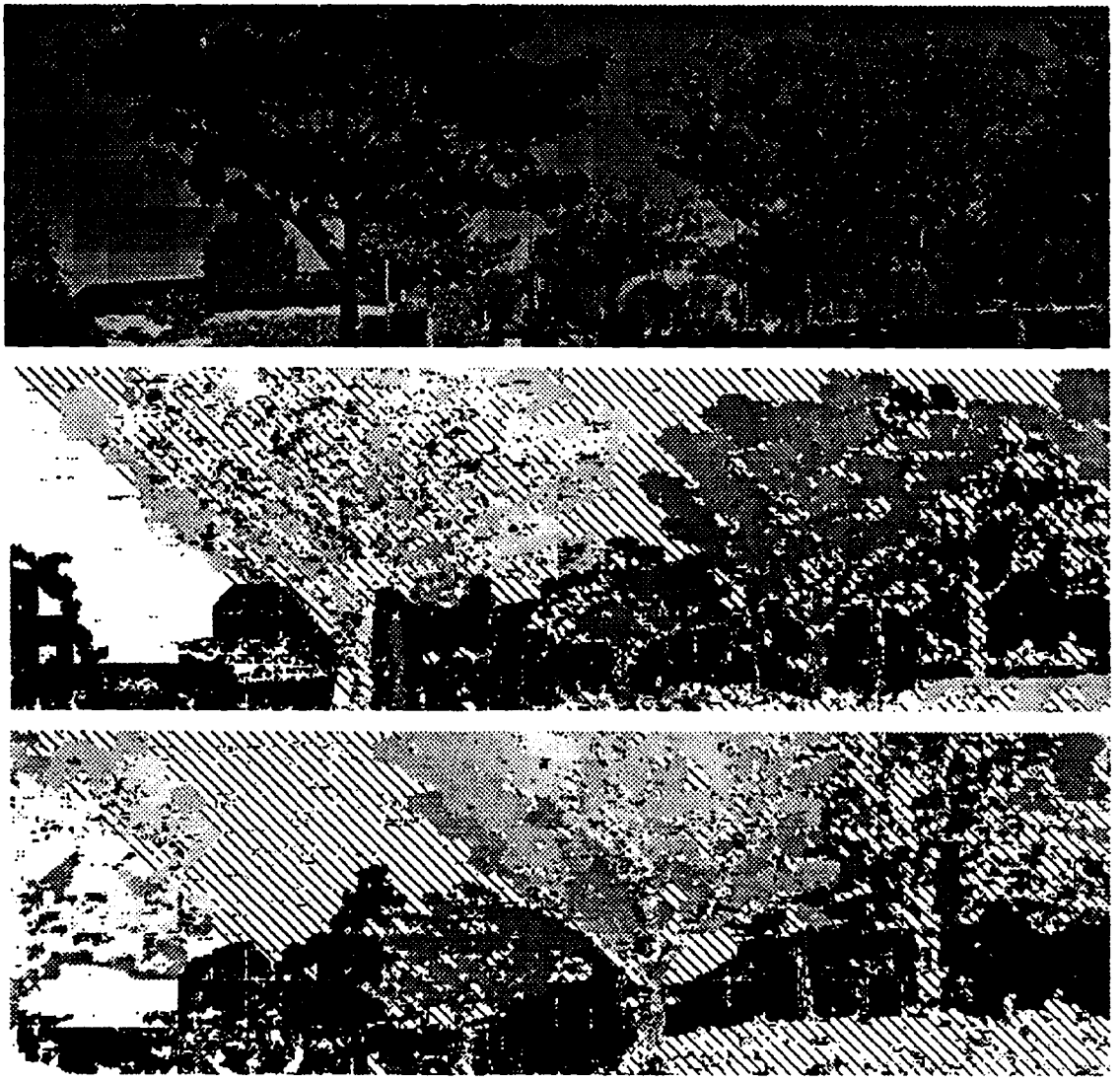


Figure 19: WFOV range maps computed from nearby locations can be used to compute local vehicle displacement. The top image here is a mosaic gray level image showing a scene covering a pan range of approximately 127 degrees. The middle image is the corresponding range map and the bottom image is a range map computed after moving the sensor 12 feet to the right. This is in the direction of the right edge of the display. The direction of the largest tree in the middle display (top-left of center) is at right angles to the direction of motion. It is half off the image on the left side of the lower display.

In assessing the usability of various analogue recording techniques, signal degradation measurements were made by taking test patterns stored in a frame buffer and recording them on video tape. These recorded images were then played back and digitized back into a frame buffer. The resulting digital images were then compared with the original test patterns to assess the degree of resolution loss going through video tape storage. These experiments indicated that a loss of a factor of two or three in horizontal resolution is incurred. While this is not an ideal situation, the restored images should be adequate for initial studies working with live video. We also looked at the spatial resolution of and crosstalk between red, green, and blue video channels decoded from VCR tape. As expected, the cross talk was much too large to be of use for recording stereo video.

We visited Robotic Systems Technology (RST) in Hampstead, Maryland along with Bob Bolles and Larry Matthies to collect video data from the stereo-intensified cameras that they have on their Surrogate Teleoperated Vehicle. RST gave us access to their equipment and had one of their staff members stay with us until 11 p.m. collecting data in the field using a pickup truck rigged with a portable generator, cameras and recorders. We were able to collect test data which will be used in evaluating our ability to perform stereo matching on intensified camera imagery.

These intensified cameras have very high shot noise levels when operating in starlight conditions, and in this case, their independent automatic gain controls occasionally caused the two cameras to produce images that were negatives of each other due to one seeing a bright object, e.g. a street light, in the distance before the other. We have done some preliminary checks on the data with our Laplacian-of-Gaussian convolver hardware to see how much stable texture could be picked out through the camera noise. It appears to be a challenging task in total darkness, but we may be able to achieve usable results with very large operators and convolution windows.

4.6 Car mounted stereo testbed facility

We have prepared our real-time facility for use on a mobile platform. Locally our stereo camera head will be mounted on a car roof-rack. Data will be collected and live processing runs will be made driving the vehicle on local back roads with topographical features similar to those at the Demo A site in Denver.

4.7 Active head control for motion tracking

We have developed a real-time image motion measurement module on our PRISM-3 system that is able to track image velocities as large as 50 pixels per frame. This module has been integrated with our active sensor head to allow following of moving objects. We will use this capability to study problems in tracking and image stabilization.

4.8 Collaboration and technology transfer

In May 1992, Teleos and SRI hosted a UGV stereo review meeting attended by stereo researchers from Teleos, SRI, and JPL along with Connie Gray from the U.S. Army Topographic Engineering Center. At the meeting we presented highlights of each group's approaches; we examined some of the stereo video test data collected thus far, including a look at this data through our Laplacian-of-Gaussian convolution hardware; we reviewed some of the results from our (static) stereo evaluation project; and we discussed how best to carry out technology transfer with Martin Marietta. This last discussion led to an e-mail document which was iterated between the three stereo contractors and then sent to Martin for their comments.

In June 1992, Teleos hosted a meeting with the UGV stereo contractors and Dave Morgenthaler and Dave Anhalt of Martin Marietta to discuss the design of the stereo hardware systems on the first UGV vehicle being assembled by Martin. SRI, JPL, and Teleos reviewed their ongoing research in stereo and objectives for the program. Discussions ensued about Martin's proposed stereo hardware architecture vis-a-vis the imaging and processing requirements we anticipate for accomplishing the UGV stereo sensing mission. The meeting was productive and set the stage for a good working relationship between the four groups.

In September 1992, Teleos prepared for and attended the NIST sponsored Workshop on Performance Evaluation of UGV Technology. The meeting yielded a productive discussion of the issues and challenges of measuring performance in the development of UGV technology.

In October 1992, Teleos hosted David Anhalt from Martin Marietta for a day to review the designs of our stereo algorithm and our accelerator hardware in preparation for the UGV workshop held during October in Denver, CO.

In addition, Teleos has actively participated in all of the UGV workshops held during 1992.

4.9 Self-narrating processes

We developed a proposal for improving user and observer awareness of the internal activity going on in our UGV systems. The proposal was accepted and will be implemented on the UGV Operator Control Unit (OCU) by Hughes for experimentation. The basic idea was to provide speech synthesizers on the various system modules that would provide "self-narration" during the execution of the demonstration. These might be on different audio channels so that the observer can switch back and forth, or, with inspiration, we might be able to have a number of speakers on a single channel. The messages would generally be canned "printf" statements with enough parameters filled-in to give a running account of what is going on in a given module. The attraction of this idea is that it doesn't take up the operator's visual attention and could lend a fast-paced feel to our demos, which might otherwise appear to be running in slow motion, especially from a distance.

5 Conclusion

This report reviewed the work done during 1992 at Teleos Research in support of DARPA's UGV program. We have discussed three aspects of our program, section two gave a broad overview of Teleos' approach to studying visual perception was presented. In it the concept of minimal-meaningful-measurement tools for early vision was described as a natural methodology for allowing a higher level application process to easily influence and exploit basic measurement modalities. Key to this were the ideas of (1) defining early measurement problems in a minimalist way so that only as much as is necessary to answer basic useful question is computed; (2) structuring the measurement module to have an easy to model behavior so that a user is better able to exploit it in new situations without having to understand the details of the internal algorithm; and (3) providing richer information about that minimal measurement, for example correlation peak shape and height along with the disparity of the peak center. The sign-correlation algorithm under development at Teleos was described in this context and the current

performance benchmarks of our accelerator technology were presented.

Section three then discussed our core research program in stereo vision. Highlights of this year's work include the development of several new algorithms for enhancing stereo matcher performance on UGV relevant imagery. In particular, Teleos developed: (1) techniques for automatically setting stereo matcher operating parameters such as filter size by previewing results on a sparse set of points over a range of possible parameter settings; (2) an analysis identifying the principal parameters affecting the magnitude of the disparity gradient effect that compromises correlator performance in UGV stereo imagery; and (3) a technique for improving area correlator performance in the presence of large stereo disparity gradients.

An equally important component of our core research effort has been in developing methodologies and tests for evaluating the performance of our algorithms in realistic contexts. To this end, Teleos collaborated with the other UGV stereo contractors on a broad evaluation of stereo matching algorithms. A suite of stereo imagery for testing matcher performance in the presence of increasing noise and in the presence of large disparity gradients was contributed to the project. Teleos also submitted its own stereo matching algorithm to the evaluation process and that algorithm performed well overall and was noteworthy among all compared for its noise handling capabilities.

The fourth section then described activities that were directed at relating the core research results to specific UGV applications. Studies were carried out in the areas of narrow-field-of-view stereo; wide field-of-view, high-resolution stereo-mosaic building; the feasibility of using stereo landmarks in support of vehicle navigation; active control of a stereo sensor head for 3-D tracking moving objects; feasibility of porting our matching algorithms to several types of parallel processors; and development of a vehicle based test facility for supporting the study of real-time stereo using an active sensor head. A significant effort was also made to support technology transfer to the UGV system integrator, Martin Marietta, and to other collaborating UGV contractors.

In conclusion, 1992 has been a productive year for us, and much ground work has been laid for continuing work on the UGV program. In the coming year we will be moving our real-time system on a vehicle and carrying out experiments in support of navigation and obstacle avoidance. We will also be working closely with SRI and JPL on integrating our ideas and transferring

results to the UGV system integrator.

References

[1] D. Marr and T. Poggio. A computational theory of human stereo vision. *Proceedings of the Royal Society of London*, 204:301–328, 1979.

[2] H. K. Nishihara. Hidden information in transparent stereograms. In *Proc. of the Twenty-First Asilomar Conf. on Signals, Systems, and Computers*, pages 695–700, Pacific Grove, CA, Nov 1987.

[3] H. K. Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, October 1984. Also in Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, edited by M.A. Fischler and O. Firschein, Morgan Kaufmann, Los Altos, 1987.

[4] J. B. Burns and H. K. Nishihara. Feasibility of using stereo to assist navigation. Technical Report TR-92-02, Teleos Research, 576 Middlefield Road, Palo Alto, CA 94301, 1992.